DOCUMENT RESUME

ED 320 122 CS 010 117

AUTHOR Valencia, Sheila; And Others
TITLE Four Measures of Topical Knowledge: A Study of Construct Validity. Technical Report No. 501.
INSTITUTION Bolt, Beranek and Newman, Inc., Cambridge, Mass.; Illinois Univ., Urbana. Center for the Study of Reading.
SPONS AGENCY Office of Educational Research and Improvement (ED), Washington, DC.
PUB DATE Jun 90
CONTRACT G0087-C1001-90
NOTE 47p.
PUB TYPE Reports - Research/Technical (143)

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS *Construct Validity; Correlation; Educational Research; Elementary Education; Grade 3; Grade 4; Grade 9; Interviews; Junior High Schools; *Knowledge Level; *Testing
IDENTIFIERS *Knowledge

ABSTRACT

A study was conducted to establish the construct validity of measures designed to assess topical knowledge. Thirty-one ninth-grade students and 44 third/fourth-grade students were interviewed to ascertain their knowledge about 4 topics; 2 weeks later they responded to 3 tests of topical knowledge. Results were analyzed quantitatively and qualitatively. Correlations between the interview scores and the paper-and-pencil measures did not reveal a strong and clear relationship between the students' performance on the interviews and their performance on the tests of topical knowledge. Conditional probability analysis revealed that the interview and the paper-and-pencil tests provided different information regarding an individual's knowledge of a topic. These differences seem to be related to the difference between a recall and a recognition task. Results indicated that the interview approach captured individual differences more clearly and more dramatically than the paper-and-pencil tests of topical knowledge. (Fourteen tables of data and four figures are included. Forty-eight references are attached.) (MG)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED320122

CENTER FOR THE STUDY OF READING

Technical Report No. 501

FOUR MEASURES OF TOPICAL KNOWLEDGE: A STUDY OF CONSTRUCT VALIDITY

Sheila W. Valencia
University of Washington

Anne C. Stallman
Michelle Commeyras
P. David Pearson
University of Illinois at Urbana-Champaign

Douglas K. Hartman
University of Pittsburgh

June 1990

University of Illinois at Urbana-Champaign
51 Gerty Drive
Champaign, Illinois 61820

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

D. ANDERSON

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☐ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

The work upon which this publication was based was supported in part by the Office of Educational Research and Improvement under Cooperative Agreement No. G0087-C1001-90 with the Reading Research and Education Center. The publication does not necessarily reflect the views of the agency supporting the research.

CS010117

BEST COPY AVAILABLE

EDITORIAL ADVISORY BOARD
1989-90

James Armstrong
Linda Asmussen
Gerald Arnold
Yahaya Bello
Diane Bottomley
Catherine Burnham
Candace Clark
Michelle Commeyras
John M. Consalvi
Christopher Currie
Irene-Anna Diakidoy
Barbara Hancin
Michael J. Jacobson
Jihn-Chang Jehng
Robert T. Jimenez
Bonnie M. Kerr
Paul W. Kerr
Juan Moran
Keisuke Ohtsuka
Kathy Meyer Reimer
Hua Shu
Anne Stallman
Marty Waggoner
Janelle Weinzierl
Pamela Winsor
Marsha Wise

MANAGING EDITOR
Fran Lehr

MANUSCRIPT PRODUCTION ASSISTANTS
Delores Plowman
Debra Gough

Abstract

This study was an effort to establish the construct validity of measures designed to assess topical knowledge. We began with the assumption that the best way to find out how much people know about a topic is to interview them. The interview then became the criterion for validating three paper and pencil tests of topical knowledge. Elementary and junior high school students were interviewed to ascertain their knowledge about four topics; two weeks later they responded to three tests of topical knowledge. The results were analyzed quantitatively and qualitatively. The correlations between the interview scores and the paper and pencil measures did not reveal a strong and clear relationship between the students' performance on the interviews and their performance on the tests of topical knowledge. This was an unexpected finding which led to further analyses (conditional probability analysis and case studies). The conditional probability analysis revealed that the interview and the paper and pencil tests provided different information regarding an individual's knowledge of a topic. These differences seem to be related to the difference between a recall and a recognition task. There was a high probability that if students gave information in the interviews, they would get that information correct on the paper and pencil tests (if it appeared there). Conversely, the probability was low that students would have mentioned in the interviews all, or even most, of the information that they got correct on the topical knowledge tests. We conclude from these findings that the information one gets of topical knowledge differs between interviews and paper and pencil measures. If the goal is to get the most complete picture possible regarding an individual's topical knowledge, then both interview and paper and pencil measures are necessary. If the goal is to assess only a specific body of information, then a paper and pencil measure might suffice; and, if the goal is to open a broader window on a student's knowledge, then an interview seems preferable. Looking across the few select cases that we analyzed in depth, we found that the interview approach captured individual differences more clearly and dramatically than did the paper and pencil tests of topical knowledge.

FOUR MEASURES OF TOPICAL KNOWLEDGE: A STUDY OF CONSTRUCT VALIDITY

In their daily routines teachers have acknowledged for years the importance of topical knowledge for reading comprehension by teaching vocabulary, building background for a selection, and even setting purposes for reading a particular text. But it took an event as major as the cognitive revolution in the psychology of learning to convince researchers to begin to study the knowledge-comprehension relationship. Between 1970 and the present, numerous studies have demonstrated that the experiences and knowledge a person brings to a text influence how he or she comprehends and recalls the text (e.g., Anderson, Reynolds, Schallert, & Goetz, 1977; Bransford & Johnson, 1973; Dooling & Lachman, 1971; Pichert & Anderson, 1977). The finding that topical knowledge affects comprehension was a natural bridge to lines of research attempting to calibrate more subtle aspects of the topical knowledge-comprehension relationship (e.g., Alvermann & Hynd, 1987; Alvermann, Smith, & Readence, 1985; Anderson & Smith, 1984; Callahan & Drum, 1984; Chou Hare, 1982; Chou Hare & Devine, 1983; Davey & Kapinus, 1985; Domaracki, 1984; Langer, 1984; Langer & Nicolich, 1981; Lipson, 1982; Maria & Blustein, 1986; Marr & Gormley, 1982; Stevens, 1980). An important outgrowth of these attempts to examine this relationship in more complex ways is that researchers have had to focus their energies on developing valid and reliable measures of readers' knowledge of text topics. In short, we have moved from broad definitions of topical knowledge and a general acceptance of the topical knowledge-comprehension connection to more specific research questions regarding the breadth and depth of that knowledge.

In many of the early studies, topical knowledge tended to be manipulated rather than measured. For example, subjects would be given texts with vague and/or ambiguous terms in order to see what knowledge they voluntarily used to make sense of them (Anderson et al., 1977). Other researchers examined the effect of disambiguating pictures or titles on comprehension of these vague passages; allegedly, the titles or pictures served as bridges to subjects' prior knowledge (Bransford & Johnson, 1973; Dooling & Lachman, 1971). Later, more direct measures of topical knowledge were used in order to try to quantify the knowledge-comprehension relationship (Davey & Kapinus, 1985; Domaracki, 1984; Langer, 1984). These studies included investigations of the breadth and depth of knowledge, and misconceptions that might detrimentally affect comprehension. Although a variety of formats have been used in these studies (e.g., vocabulary, multiple-choice, free response), few have been accompanied by any documentation of efforts to evaluate their validity and reliability.

Validity and reliability are important characteristics of any measure from which educators wish to draw inferences about either individuals or groups. Given the complexity of the topical knowledge construct (see, for example, Spiro, Feltovich, Coulson, & Anderson, in press), the diversity of the measures used across studies, and the widespread application of the topical knowledge-comprehension research findings, the need for validity studies is clear. If, in fact, measures that have been used to assess topical knowledge are not valid, then the assumptions and conclusions that have been based on this body of research are brought into question. Some scholars (e.g., Johnston, 1984a; Messick, 1981) have argued that construct validity is one of the most important concerns in assessment. In this study, we began with our best definition of the construct and then tried to determine what aspects of topical knowledge could be assessed by each of several candidate measures.

The studies of most interest to us are those that have attempted to measure topical knowledge directly. Within these studies, the types of measures, the scoring systems, the kinds of topical knowledge assessed, and the types of texts vary considerably. These differences demand critical evaluation of various operational definitions of topical knowledge and the ways in which these implicit definitions have led to theories regarding the role of topical knowledge in comprehension. And thus, it is these differences that serve as the framework for our discussion of various topical knowledge assessments.

Formats for Measuring Topical Knowledge

Researchers have used at least seven different formats for assessing topical knowledge. The formats range from those with a recognition orientation, such as multiple-choice questions, to those with a recall orientation, such as written statements of what an individual "knows" about a topic. We have placed these various formats on a continuum that illustrates the type of demand the assessment task places on the student (see Figure 1). Recognition tasks reflect the body of knowledge that the researcher has decided is critical to the comprehension task at hand. The respondents' task is to react to the stimulus by choosing an answer from a set of items. In contrast, the recall tasks are more open-ended and provide greater latitude for those who are asked to demonstrate their knowledge. On the recall end of the continuum, typically students are presented with some general topic label or situation and asked to tell or write what they remember about the topic. In between the multiple-choice knowledge test and the "what do you know about X" recall task are many variations each depicted in Figure 1 and discussed below.

[Insert Figure 1 about here]

Multiple-choice content questions. This type of topical knowledge measure was used in 6 of the 19 studies (see Table 1). These questions require students to recognize correct information about the content of the specified topic. Alvermann, Smith, and Readence (1985) developed a 20-item multiple-choice test with one correct answer and two distractors per item. The following sample typifies multiple-choice recognition assessment of topical knowledge:

Which of the following is true about rattlesnakes?

a. Rattlesnakes hide for protection.

b. Rattlesnakes chase people.

c. Rattlesnakes can strike over 3 times their length.

[Insert Table 1 about here]

While some studies use three or even four distractors, the basic multiple-choice recognition format is the same. It is not surprising that these types of questions constitute the most frequently used type of topical knowledge measure. They are also the most common question type used for assessing comprehension in instructional materials (Armbruster & Ostertag, 1989; Foertsch & Pearson, 1987).

Multiple-choice vocabulary tests. This type of topical knowledge measure was used in two studies. Typical of these vocabulary tests, Johnston (1984b) gave subjects 33 content-specific multiple-choice questions, each presenting a word and four possible definitions or a definition and four possible words. For example:

POLLEN

a. seeds

b. male germ cells from anthers

c. part of the steering gear on a ship

d. a type of bee

Again, it is not surprising that vocabulary tests have been used to assess topical knowledge. They have been used successfully to measure other constructs, such as IQ and achievement. And, more importantly, vocabulary knowledge is often viewed as a window on an individual's knowledge about a given topic (Anderson & Freebody, 1983; Johnson & Pearson, 1978, 1984).

Structured word associations. A type of structured word association task was used only by Domaracki (1984). All possible combinations of eight key concepts drawn from a passage were randomly paired into relatedness judgment tasks; each pair of key concept words was followed by a 7-point scale anchored by the words "highly related" at one end and "not related" at the other. Although Domaracki did not provide a sample item, an example of a structured word association might well look like this:

snake : radio

!_____!_____!_____!_____!_____!_____!_____!

highly related not related

This format is a cross between a semantic decision task and a Likert scale, requiring a judgment about the relationship between the two concepts. Domaracki considers this task a measure of the depth of knowledge held by the student rather than a measure of the breadth of topical knowledge.

Completion questions. Two studies (Chiesi, Spilich, & Voss, 1979; Spilich, Vesonder, Chiesi, & Voss, 1979) used completion questions to assess topical knowledge of the domain of baseball. Although the exact form of the completion questions was not specified in either article (nor were examples provided), these items were most likely similar to those in which students must fill in the blank either in the middle or at the end of a sentence.

The batter hit the ball and ran for _____ base as fast as he could.

Or

You would steal home if _____.

Structured or direct questioning. Structured questioning of topical knowledge measures fall into two subcategories: those in which students were interviewed and those that required written responses. Typical of the five studies using structured or direct questioning in interviews is Holmes' (1983) work. For example, on the topic of snakes' skin, she asked, "Does a snake keep the same skin its whole life?" The only study requiring written responses to structured or direct questioning was that of Davey and Kapinus (1985). Their questions asked the students to write as much as they could in response to eight questions concerning the location, function, appearance, and workings of computers.

Written and oral free word association. One format for assessing breadth and depth of topical knowledge is a written open-ended response to word associations (Langer, 1980, 1984; Langer & Nicolich, 1981; Chou Hare, 1982). Typically the experimenter selects key words or phrases to represent the major concepts of a passage. Students are then asked to free associate in response to each concept--to jot down anything that comes to mind when they hear that particular word or phrase. Responses are judged according to a content hierarchy of ideas. Findings from all five studies using written free word associations (Callahan & Drum, 1984; Chou Hare, 1982; Langer, 1980; Langer, 1984; Langer & Nicolich, 1981) indicate that this type of topical knowledge assessment is strongly related to subsequent recall of passages.

Oral free recall. This type of task represents the least directive and least structured task, and thus perhaps the most "natural" and least likely to be influenced by reading or writing ability. Holmes and Roser (1987) used two types of free recall--one in which students were to tell as much as they knew

about a given topic and the second in which they were to talk about their first-hand experiences related to the topic.

This review of formats represents the majority of measures used to assess topical knowledge and clearly demonstrates the variation in measures. This diversity is best seen by their spread along the recognition-recall continuum in Figure 1. With such variation in the instruments and the procedures used to assess topical knowledge, it is important to evaluate the validity of these measures; are they measures of the very same construct, or are they multiple measures of slightly different facets of the same global construct, or are they measures of totally independent constructs? Current research leaves these questions unanswered.

Types of Scoring

With respect to scoring, the studies fell into two general groups: those that focused on the quantity of ideas (breadth of topical knowledge) and those that focused on the quality of those ideas (depth of topical knowledge). Most of the scoring systems were quantitative (16 out of the 19), counting the number of correct or related ideas as an indicator of how much a student knew. Four of the six qualitative scoring systems were based on the work of Langer (1980, 1984), who examined the nature of the hierarchical relationships among the concepts offered in free association tasks (e.g., collie is an example of dog, whereas bark is an attribute).

Interestingly, three studies used both quantitative and qualitative scoring systems. Chou Hare (1982) scored students' free associations about key words both quantitatively (total number of words) and qualitatively to find out which scoring of topical knowledge better predicted comprehension. Her results indicated that both scoring systems predicted overall passage recall, but that quantitatively scored topical knowledge was a better predictor. In contrast, Domaracki (1984) used a vocabulary test (scored quantitatively) and a concept relatedness task (scored qualitatively) to investigate the relationship between the amount and structure of topical knowledge to reading comprehension. Her results showed that the qualitatively scored measure better predicted total comprehension and that it was redundant with ability measures. Holmes (1983) scored student responses to structured and direct questions both quantitatively and qualitatively (i.e., accuracy of ideas). Her results were based solely on the qualitative scoring; they indicated that poor readers and good readers were comparable at answering literal comprehension questions related to information they knew before reading, but poor readers were not as adept as good readers in using their topical knowledge to answer inferential comprehension questions.

Types of Text

Another factor of interest in the topical knowledge literature is that of text type. There are clear differences between the structure of narrative and expository texts (Mandler, 1984; Olson, 1985) and, quite obviously, there are differences in the content and conceptual load. Such differences are likely to require parallel differences in the underlying cognitive processes necessary to comprehend each. Thus, it is interesting to consider the implications for the measurement of topical knowledge of narrative and expository texts.

By far, most studies assessed topical knowledge and comprehension for expository text (16 out of the 19); only one study (Chou Hare & Devine, 1983) used narrative texts, and two used what might be called seminarrative texts (see the description of the goal structure and content of "narrative" baseball passages in Spilich, Vesonder, Chiesi, & Voss, 1979; Chiesi, Spilich, & Voss, 1979). Others, such as Graves, Cooke, and LaBerge (1983), have looked at the effects of previewing (a topical knowledge activation strategy) on difficult narrative texts, but have not directly addressed the assessment of topical knowledge for narratives. With so few studies on topical knowledge for narrative texts, it is not clear whether genre differences exist.

Type of Prior Knowledge Assessed

The measures used in most studies can be classified as assessing two types of knowledge: content knowledge and knowledge of the structure or genre. The former type of knowledge refers to the actual information the student possesses and the latter to the organization of that knowledge. Most of the studies looked solely at the content of topical knowledge (14 out of the 19). The remaining five studies looked at the structural topical knowledge combined with the content of topical knowledge. Langer (1980, 1984), Langer and Nicolich (1981), and Chou Hare (1982) looked at the content and structure of written free word associations, while Domaracki (1984) looked at the content in a vocabulary test and structure through a structured word association task.

Age and Ability of Subjects

The concern here was with the composition of the student populations employed in topical knowledge research. Subjects in these studies ranged from first grade through graduate school; ten studies used students from the elementary grades (1-6), three from the middle grades (7-8), four from high school (9-12), and two from higher education (undergraduate and graduate). Although several studies distinguished good readers from poor readers, none attempted to discern if topical knowledge operated differentially for students of different developmental levels.

Validation of Measures

Clearly, a number of measures and operational definitions have been used to assess topical knowledge. However, few researchers have attempted to establish the validity of their instruments; that is, what assurance is there that the various instruments are measuring a common construct called "topical knowledge"? Only 5 of the 19 researchers whose studies we examined had attempted to validate the topical knowledge measures themselves. The unstated assumption in most of the studies is consistent with the rationale provided for the validity of free word association measures: They are "presumed valid because both the concept and the vocabulary knowledge revealed by such measures are implicated in comprehension" (Chou Hare & Devine, 1983, p. 1). Such reasoning can be construed as circular.

What little work has been done can be clustered into two categories: (a) those studies that distinguish topical knowledge from other, perhaps competing, constructs such as general ability or interest, and (b) those that attempted to establish concurrent validity with other measures of topical knowledge. Johnston (1984b) addressed the validity question indirectly by demonstrating that content specific topical knowledge and general ability (IQ) make independent contributions to reading comprehension. Similarly, a few other researchers have attempted to distinguish topical knowledge from interest (Chou Hare & Devine, 1983; Baldwin, Peleg-Bruckner, & McClintock, 1985). These studies have suggested that topical interest and topical knowledge were autonomous factors in reading comprehension, in fact, the former study indicated that interest did not predict comprehension.

The second approach to validation is represented in the work of Holmes and Roser (1987) and Chou Hare (1982), in which at least two different forms of topical knowledge assessment were compared. In both cases the focus was on which measure was the most effective and efficient. Chou Hare found that for word associations, a quantitative scoring scheme predicted comprehension better than did a qualitative scheme. Holmes and Roser (1987) concluded that structured questions produce the greatest quantity of facts as compared with more open-ended approaches. Interestingly, both studies prioritized quantity either in the form of "best predictor" or in the form of "most information." Neither looked at the differential information contributed by each assessment technique.

Domaracki (1984) suggests that all topical knowledge measures may not contribute in the same way. In a comparison of structured word associations and multiple-choice recognition questions, she found differential predictive power of quantity of topical knowledge and quality of topical knowledge. Thus,

amount (breadth) and structure (depth) of knowledge each explained a unique proportion of the total reading comprehension performance. When the focus is not on which measure is a better predictor, there is an indication that different measures may be tapping different aspects of topical knowledge.

The Current Study

It is against this backdrop--the wide differences in the nature and structure of topical knowledge measures--that we developed a set of questions to guide our research. The first question pertains to the validity of the measures: Do various paper and pencil assessment devices developed by teachers and researchers actually measure topical knowledge? To answer this question we began with an important assumption that the very best way to learn what people know about a topic is to encourage them to tell you what they know about that topic. This assumption provided an important starting point for defining the topical knowledge construct. Rather than define the topical knowledge construct by its relation with comprehension and thus accept all previous measures of topical knowledge as valid, we chose to begin with the definition of the construct as operationalized in the interview. We assumed that a relaxed interview situation in which the interviewer could follow-up on promising leads would elicit the best possible estimate of a student's knowledge about a topic. Then, given several paper and pencil measures constructed to assess the same areas of topical knowledge as the interviews, we would be able to evaluate how well each of these written measures aligned with the ideal interview approach. Although this seemed like a reasonable beginning point, we were aware that, based on our results, this assumption would need to be reexamined throughout the study.

The second question addresses the generalizability of any findings emerging from any attempt to answer the first question: Is the measurement of topical knowledge conditioned by factors such as type of measure, genre (expositions versus narratives), and student level of maturity (in our case, third-grade versus ninth-grade students)?

Method

Subjects

Subjects for this study were junior high school and elementary school students from a small midwestern city. In all, 31 ninth-grade students and 44 third/fourth-grade students participated in this study. None of the students had been identified as having any educational disabilities or limited proficiency in English. Results from the reading section of the Comprehensive Test of Basic Skills (CTB McGraw-Hill, 1983) indicated that these groups of students were reading at or near their grade level placements.

Measures of Topical Knowledge

Students' knowledge on four topics was assessed using four different measures of topical knowledge. Three of the measures were pencil and paper activities and the fourth was a structured interview. In all, six topics were used in this study, three derived from narrative texts and three from expository texts. The third- and fourth-grade students were assessed on their knowledge of a boy's first day at a new school (New Boy), and how plants and people help each other (Plants and People). The ninth-grade students were assessed on their knowledge of a girl who spies to protect a well-known man (Spy) and human blood circulation (Human Circulation). Both the third/fourth- and ninth-grade students were also assessed on the topic of a girl who baby-sits for other people's pets (Petsitter) and animal defense mechanisms (Animal Defenses). All the texts from which these topics were derived were taken from materials students typically read in school (basals, textbooks, student magazines); thus the topics were judged to be ecologically valid. The measures were piloted and items were revised based on the pilot data.

Vocabulary measure. One of the measures developed was a multiple-choice Vocabulary test. As in many of the topical knowledge studies, words were selected that represented the key concepts and

ideas for each topic. While some of these words might have appeared directly in the source texts, others were not presented in the texts but were representative of key concepts related to the topic itself. Following a traditional format, 10 words for each topic were presented in isolation with instructions to select the best definition from among five choices.

Two tests were developed--one for ninth-grade students and another for the third- and fourth-grade students. Each test assessed students' knowledge of the 40 (10 per passage) randomly ordered words. Four scores, one for each passage at each grade level, were calculated for the Vocabulary test. Students received 10 points for each correct item to obtain scores ranging from 10 to 100. For example, a score of 90 on the test for Animal Defenses means the student knew 9 of the 10 words associated with that topic.

Although previous studies have used vocabulary measures as indices of topical knowledge, they have rarely applied the format to narrative topics. The construction of a vocabulary measure for narratives posed some interesting problems. The identification of key concepts for narrative passages was more difficult than for expository passages because the world of possible relevant concepts is less constrained in narratives. For example, certain words (*heart*, *capillaries*, and *veins*) are quite obviously key concepts related to topical knowledge about human blood circulation. In contrast, narrative topics such as a girl who baby-sits for pets, elicits a wide range of relevant concepts. Words like *resourceful*, *conscientious*, *obligation*, and *profit* seem important, but they could be easily replaced by a set consisting of *satisfaction*, *entrepreneur*, and *advertising*. Furthermore, it is not difficult to imagine other narratives for which all of these words could also be used to assess topical knowledge.

Circle measure. This measure was similar to a vocabulary measure and was one of several that was developed for the Illinois Assessment of Reading (Valencia & Pearson, 1987). Students were presented with a word or phrase representing a major topic and 24 words representing a range of concepts related to the topic (see Figure 2). For each word, students had to decide whether it was highly related to the topic, somewhat related, or not at all related.

[Insert Figure 2 about here.]

The purpose of this measure was to assess students' knowledge of more or less hierarchical relations between many of the concepts and the focal topic. This was visually presented using three large concentric circles and a small rectangle placed in the upper left hand corner of the page. This provided a visual metaphor for semantic relatedness between the words and the topic. This measure was only used to assess topical knowledge for expository passages. As with the Vocabulary measure, there were limitations to using this procedure for narrative topics. Because it was difficult to imagine words that are not at all related to the global themes of most narratives, the Circle measure was not used to assess topical knowledge for narrative topics.

The items on the Circle measure were developed using a model of developing expertise. For each topic, an item generation matrix was constructed representing two continua: level of expertise (novice, average, expert), and level of relatedness to the topic (highly related, somewhat related, not at all related). The resulting 3 x 3 matrix provided the framework for item construction for this task.

The items on the Circle measure were scored using a "discrepancy from expert" model. First, as a check on the item generation matrix, an expert answer key was developed by administering each test to three groups of experts (the test designers, adults, and students from a higher grade level). Validation using all three groups was required before a response was assigned to a relatedness category (i.e., highly related, somewhat related, or not at all related). That is, all three groups had to agree on the classification of the word, or the more sophisticated groups of experts had to show an increasing tendency to place it in a given category. For example, *camouflage* was categorized as highly related to animal defenses by all three expert groups and therefore keyed as highly related. But for the word *predator*, the percentage of "experts" judging it as highly related increased with the presumed "expertise"

of the expert groups. So, for example, 60% of the older students chose it as highly related, while 85% of the adults and 100% of the test constructors selected it as highly related. Discrepancies and inconsistent results were identified during piloting and resolved before the administration of the final version.

Second, student responses were assigned points on the basis of their degree of agreement with the expert key. Therefore, if a student agreed with the expert answer key, the score received was a 1; if the student's response only partially agreed with the experts (e.g., the student selected "highly related" when the keyed response is "somewhat related"), then the score was .5, and if there was a wide discrepancy between the selected response and the keyed response (e.g., the student selected "not at all related" when the keyed response was "highly related"), the score was 0.

The Yes/Maybe/No measure. This pencil and paper measure of topical knowledge was also developed for the Illinois Assessment of Reading (Valencia & Pearson, 1987). Students' knowledge about a topic was assessed in response to a prompt encouraging them to think about what they knew about a specific topic, story line, or theme in the context of a particular type of selection. The inclusion of a text type reference was important to help students focus their thinking in terms of expectations for an informational piece or a narrative one. For example:

> "Pretend you are going to read a story about a girl who knows about taking care of other people's pets. Think about what you know about taking care of pets."

or

> "Pretend you are going to read a library or science book about how animals hunt other animals to get food, and how the hunted animals try to protect themselves. Think about what you know about how animals hunt one another for food and about how the hunted animals defend themselves."

Then they were asked to predict the likelihood that specific ideas might appear in such a passage. An example of a prompt and the associated items in the Yes/Maybe/No format appears in Figure 3.

[Insert Figure 3 about here.]

The students were asked to judge the likelihood of occurrence of a range of options that represent predictions generated by expert, average, and novice readers. As in the Circle format, items were generated using a matrix representing two continua: one representing expertise, and the other degree of relatedness to the topic (Yes, Maybe, or No).

Scoring for this format was also similar to that used for the Circle format. Students received 1 point for an exact match with the answer key, .5 for a partial match (e.g., selecting "Yes" when the keyed response was "Maybe"), and 0 when there was a wide discrepancy with the expert key (e.g., selecting "No" when the keyed response was "Yes"). The criteria for keying a given response as a Yes, Maybe, or No were based on the data from "expert" samples. The procedures used were virtually identical to those described for the Circle format.

The Interview. An interview was included to serve as the criterion measure against which all the different paper and pencil measures would be judged. We assumed that an interview would provide the best possible measure of a student's topical knowledge. Many researchers and clinicians have found interviewing to be a valuable method for investigating people's beliefs, attitudes, values, knowledge, and "mental content" (Gordon, 1980).

We attempted to conduct our interviews with students in a relaxed, conversational manner. Interviews as "conversations with purpose" have a long tradition in social science research (Burgess, 1984). Webb

and Webb (1932) recommended making the interview as pleasing and agreeable as possible for the persons being interviewed. Achieving this comfort depends, in part, on developing some trust and confidence between the interviewer and respondent (Oakley, 1981; Finch, 1984). In order to put the students at ease, we began by engaging them in a conversational interview about a topic of their choice.

The design of our interview procedures included aspects of scheduled and nonscheduled interviews. The interviews began with an established set of questions, which is characteristic of scheduled interviews. Gordon (1980) notes that scheduled interviews, with high topic control, are generally more efficient and effective in providing coverage, precision, and reliability of measurement. Following each established question, we probed students for additional information or clarification of unclear information. This probing is characteristic of nonscheduled interviews. Queries posed in a nonscheduled interview are intended to allow the respondents to follow their own natural paths of associations. The interviewer must be free to probe for additional details and to clarify vague points (Gordon, 1980). Our interviews consisted of three parts (A, B, C), each of which included elements of scheduled and nonscheduled interviews.

Part A of each interview began with a prompt that provided the same information that was used in the Yes/Maybe/No prompt. Thus, it led the students to think about the topic and provided information about the type of text in which this information might be found. Subjects were allowed to talk as long as they wished in response to this prompt. The interviewers set a standard 20-second wait time for determining when to proceed to Part B of the interview.

In Part B students were asked questions that permitted decomposition of the information in the initial prompt into more specific and concrete topics in order to try to get some information from even the most reluctant students. For example, Part B of the petsitting topic began with the prompt, "Tell me what you know about taking care of pets." When students finished responding to that prompt, they were asked, "Pretend you have a petsitting service. What kinds of things do you think might happen?" All students were given both parts A and B of the interview.

For the most recalcitrant of students, there was a part C. It was given only to those students who had failed to give any responses to the prompts used in parts A and B. Questions in part C focused on personal experiences students may have had related to the topic. These questions helped scaffold students' responses to bring them to a point where they might be able to verbalize their knowledge about the topic. For example, questions about the petsitting topic included:

1. Do you have a pet? How do you take care of it?

2. What does your pet like to do?

3. If you could have a pet, what kind would you want? Why?

4. If you were going to have somebody take care of your pet, what would you want them to do?

Embedded within this overall structure of interview probes were procedures for clarifying ambiguities, eliciting additional information, and checking on seemingly irrelevant information. Interviewers were instructed to seek reasons behind the information students provided. These procedures were used to tap students' knowledge more deeply. A response to a probe typically was followed by a question like, "Tell me more about taking care of pets." or "Why do you think that is true?"

The process of scoring the information from the interviews involved a series of steps. The first step was to create a template for each of the six topics. The templates represented a complete and ideal knowledge base for students in the study; it was derived by compiling all of the information collected for all students interviewed and by reviewing the information that potentially could be expected from

students at that level. The second step was to use the template to categorize all the ideas a particular student gave in the interview. Three raters listened to audio tapes of the interviews and recorded which ideas from the template appeared in each part (A, B, or C) of each interview. Misinformation was also noted. Using this procedure, the average interrater reliability among the three raters was 92%.

The third step was to organize the ideas on each template according to three levels of conceptual complexity that had been determined a priori. Level 3 ideas were superordinate concepts, Level 2 ideas were subordinate concepts, and Level 1 ideas were related details. For example, for the animal defenses topic, predator and camouflage were judged to be Level 3 ideas, while examples of predator/prey relationships, such as bird and worm, were judged as Level 2 ideas. Ideas conceptually related to Level 2 ideas, such as seasons or nests, were categorized as Level 1 ideas. Additionally, a separate tally was kept for all ideas that represented misinformation or unrelated information.

The fourth and final step was to use the key built in the third step to categorize the ideas given by each student for a given topic. At a later stage, categories were converted into points to create "conceptual richness" scores. Those responses at the highest level of conceptual richness, Level 3, received a score of 3; those at the next level, Level 2, received a 2; and those most subordinate were scored 1. No credit was given for unrelated or redundant ideas, nor was credit subtracted for misinformation. Thus, both the conceptual level of each idea and the stage (Part A, B, or C) at which it occurred were recorded for each unique idea for each student. The interrater reliability for these classifications was 96%.

Procedure

Students were interviewed individually on two occasions. At each session, the students were interviewed by a different examiner about two randomly ordered topics. Thus, interviewer and passage effects were randomized for the interviews. Interviewers took extensive notes on students' responses using a specially prepared checklist, and they tape recorded each interview for subsequent reliability checks. The interview sessions lasted from fifteen to twenty-five minutes depending on how much each student had to say and the extent of the probing needed.

To minimize carry-over effects from the interviews, students were given the pencil and paper measures of topic familiarity (Vocabulary, Circle, and Yes/Maybe/No) two weeks after the second set of interviews was completed. The researchers read all the pencil and paper items aloud to groups of students so that decoding ability would not be confounded with the topic familiarity assessment. The three pencil and paper measures were randomly ordered for each of four small group administrations. Tests were administered in two sessions, one for the Vocabulary test and the other for the Circle and Yes/Maybe/No tests.

Results

Descriptive Data

Tables 2 and 3 present the mean, standard deviation, and range for each prior knowledge measure by grade level. The means of the measures indicate that there were pronounced differences both across and within text types for all three paper and pencil measures and for the interviews. In order to compare them more directly, the scores for the paper and pencil measures were all converted to a scale from 0 to 100. The interview scores represent the total number of points the student received for the ideas given in the interview. The ranges of the scores on all the measures show that there was variability among the students and their knowledge of the topics.

[Insert Tables 2 & 3 about here.]

Correlational Analyses

To establish the construct validity of the three paper and pencil measures, we began with the assumption that the optimal strategy for determining what people know about a given topic is to provide an opportunity for them to talk about it. Hence it became the criterion against which all other measures were "evaluated." It was assumed that a high correlation between a paper and pencil measure and the interview would provide evidence for the validity of that measure by showing that it captured the same level or type of topical knowledge that was demonstrated in the interviews. Therefore, the first step in the analysis of the data was to examine the relationships between paper/pencil and interview scores for each topic.

It was expected that the correlations would be high, reflecting the fact that while there are differences in students' knowledge from one topic to another, their knowledge about any given topic is fairly stable regardless of test format. However, this was not the case (see Table 4). At third grade, 4 of the 10 possible correlations with interview scores were significantly different from 0; at ninth grade, only 1 of the 10 correlations achieved significance. Not only were few of the correlations significant, there were no systematic patterns among them. We were unable to calculate the correlation between the Circle measure and the interview measure for the Plants and People passage because of the unreliability of that Circle measure. Another interesting fact revealed by these data is that the intercorrelations among the paper and pencil measures were generally not significant (r's ranged from .39 to -.23).

[Insert Table 4 about here.]

The next step was to conduct multiple regression analyses to examine the proportion of the variance in the interviews that could be explained by the students' performance on all three paper and pencil tests for each topic (see Table 5). As expected from the first order correlational analysis, the students' performance on the paper and pencil measures did not account for much of the variance in their interview scores. While more of the variance could be accounted for in the expository topics than in the narrative topics, there was wide variability between passages within a genre. Again, this finding had not been anticipated.

[Insert Table 5 about here.]

Without clear patterns emerging from the correlational and regression analyses, other explanations for results and other methods of analysis were investigated. One explanation for the unexpected findings might have been the possible unreliability of the paper and pencil tests (see Table 6).

[Insert Table 6 about here.]

These reliabilities are generally lower than standards used for commercial tests; however, they are similar to those obtained for other measures of topical knowledge (see Valencia & Pearson, 1987). Although subsequent analyses after correcting for attenuation increased the magnitude of the correlations overall, there was still little evidence of a systematic pattern in the relationship between the interviews and the paper and pencil tests either within or between topics (see Table 7). As revealed in the first series of analyses, there is a slightly greater tendency for higher correlations between tests and interviews for expository topics than for narrative topics, but this finding is not consistent. From these analyses, we concluded that poor item construction was not the entire reason for the inconsistent results.

[Insert Table 7 about here.]

Because there did not seem to be a satisfactory statistical explanation for the obtained relationship between the paper and pencil tests and the interviews, we turned to a more fine-grained qualitative analysis. A cursory examination of the kinds of information elicited from the various formats,

reinforced by the low intercorrelations among the measures, suggested the possibility that different formats might be tapping different types of information. To examine this possibility, the data were subjected to a conditional probability analysis.

Conditional Probability Analysis

This post hoc method of analysis evolved from a question about how the students performed on items that were present on at least two of measures of topical knowledge. We wondered whether a student who demonstrated knowledge about a particular idea or concept on one measure of topical knowledge would demonstrate knowledge of that same idea or concept on the other measures.

A matrix of common items was designed for each of the six topics. These matrices reflected the overlapping items between the measures of topical knowledge. For example, on the topic of animal defenses, the concepts of *camouflage*, *predator*, *carnivore*, and *scavenger* were common to the Yes/Maybe/No task and the Vocabulary test. Other concepts such as *prey* and *habitat* were common only to the Circle task and the Vocabulary test. The amount of overlap among the four measures of topical knowledge varied. For the third-grade measures there was an average of 6.6 items (range = 0-12) that overlapped between any two of the measures. For the ninth-grade measures there was an average of 5.6 items (range = 2-12) of overlap.

Determining what constituted overlap (i.e., a conceptual match) between items on the four measures of topical knowledge was much clearer for the expository topics than it was for the narrative topics. Items on the expository measures were usually exact matches, such as the words *predator* or *camouflage*, or they were very clear explanations of specific concepts. For example, an interview response offering specific examples of how animals hide from other animals would qualify as a match with the word *conceal* on a paper and pencil measure.

In contrast, a match on narrative topics required a much broader notion of equivalence. For example, in the interviews about a girl who wants to earn money by baby-sitting for other people's pets, students said she would *get customers*, *advertise*, *put up signs*, and *outline services*. It was decided that the Vocabulary test item *entrepreneur* was a concept that matched these four interview ideas. In another example, an item on the Yes/Maybe/No test, *She opens a petsitting service to make money*, was judged to match the Vocabulary item *entrepreneur* as well as entrepreneurial activities mentioned in the interviews. The rationale for such scoring is based on the priority given to the concepts related to the topics rather than the definition of prespecified words; in this case, despite the variations in wording, the underlying concept of entreprenurialship rendered these responses equivalent.

The matrices also included the ideas or concepts that appeared only once on any particular measure. This permitted us to determine the unique contribution of information elicited from each different format. The following information was recorded for students at each grade level:

1. The percentage of agreement between information given in the interview and each paper and pencil measure.

2. The percentage of agreement between correct responses on the paper and pencil measures and corresponding information from the interview template.

3. The percentage of information given in interviews that was not tested by any items on any of the paper and pencil measures.

4. The percentage of correct information from paper and pencil measures that was not given in any of the interviews. This was done separately for each paper and pencil measure.

5. The percentage of students' consistency of performance on common items from any two paper and pencil measures.

The results of this analysis (see Tables 8-11) reveal some trends that provide us with a hypothesis about prior knowledge that is conceptually more complex than the view of topical knowledge held at the inception of this study. We speak of the results of this analysis as "trends" because the standard deviations for the percentages calculated were large. Despite the large standard deviations, there are telling similarities in the overall percentages calculated on the various conditional relationships that were explored.

[Insert Tables 8-11 about here.]

The first trend, consistent across both grade levels and topics, is that the probability that information given in the interview would be correctly identified on a paper and pencil measure is higher than the probability that information students recognized on a paper and pencil measure was given in an interview. On average, students correctly recognized 76% of the items on the paper and pencil measures that they mentioned in their interviews. By contrast, only 31% of the items correctly recognized on a paper and pencil measure were ideas mentioned in their interviews. This trend reflects the difference between recognition and recall of information. In a sense, the paper and pencil measures provided students with the opportunity to display topical knowledge they had failed to voluntarily recall when they were interviewed. On the other hand, information that had been voluntarily recalled three weeks earlier was, for the most part, available during the recognition tasks.

The second trend is related to ideas that appeared on paper and pencil measures but never surfaced on the interview templates (see Table 12). For a given paper and pencil measure, unique information refers to those ideas that were assessed in the pencil and paper measures but were never mentioned by any students in the interviews. For example, the Circle task asks students to determine the degree of relatedness of the word *omnivorous* to the topic of animal defenses (predator/prey relationships), yet the idea that some animals eat both plants and animals did not surface in any of the interviews conducted with students. Thus, *omnivorous* is an idea that is uniquely assessed in one of the paper and pencil measures of topical knowledge.

Even though students never voluntarily mentioned any of the unique items, they were able to recognize, on average, 77% of these items on the paper and pencil measures. Hence, the inclusion of such concepts on the paper and pencil measures provided an index of information that was less readily accessible or was less likely to be shared in an interview.

The amount of unique information acquired from the paper and pencil measures is higher for the expository topics than it is for the narrative topics. For both third and ninth graders there is about 10% more unique information contributed by paper and pencil measures for expository topics than for narrative topics. This indicates that students have more information available with respect to expository topics than they tend to offer in an interview situation.

[Insert Table 12 about here.]

The third trend concerns the high percentage of unique information students contributed in their interviews. Much of the information given by students in the interviews was never tapped on any of the three paper and pencil measures of topical knowledge. On average, 66% of the topically relevant ideas students gave during interviews were not tested on any of the paper and pencil measures. The unique contribution of ideas revealed through interviews is considerably higher for expository topics (76%) than it is for narrative topics (56%). The point about this unique information is that interviews revealed a view of students' knowledge that would not have been evident simply by looking at their performance on paper and pencil measures. Additionally, in terms of sheer numbers, students offered

approximately 10 to 12 times the number of ideas for a particular topic than could reasonably be included in a paper and pencil measure constructed for that topic.

The fourth trend informs us about the consistency of students' performance on equivalent items across the three paper and pencil measures of topical knowledge. Consistency was examined at the individual item level, that is, when a student received the same score--correct or incorrect--on two different measures for the same concept, such as *camouflage*. On average students scored consistently on about 70% of the common items. Students of lower ability were less consistent (60%) than students of middle ability (68%), and together they were less consistent then students of high ability (74%). The variation in students' consistency may be due in part to the level of error in the scores of students of differing abilities or the novelty of the different test formats.

Four Case Studies

To gain an even more concrete picture of how these trends operate, two third-grade students and two ninth-grade students were selected for case studies. From each grade we chose one student who had performed very well on the interview as well as one who had performed poorly. Using information gathered from one topic--animal defense mechanisms--each interview was transcribed and scored against the template. Then performance was compared on the three paper and pencil measures (Vocabulary, Yes/Maybe/No, and Circle) and the interview using standardized scores. Hence, there were four comparable views of each student's knowledge.

Results suggest that students of widely differing interview performance demonstrated a much narrower range of performance on the paper and pencil tests (See Table 13). It is interesting to note that the high-interview third grader scored substantially lower than the low-interview third grader on two of the three paper and pencil measures (Vocabulary and Yes/Maybe/No). Furthermore, the high-interview ninth grader and the low-interview ninth grader both obtained similar scores on the three paper and pencil measures (within half of a standard deviation). The interview scores provide us with a very different understanding of these students' knowledge about animal defense mechanisms. In fact, using two of the three paper and pencil measures, we would be led to reclassify the "low interview" third-grade student as the "high" knowledge student. Similarly, we might conclude that there are no differences in topical knowledge for our ninth-grade students as measured by all three paper and pencil measures even though they have interview scores that differ by more than three standard deviations.

An additional interesting finding can be observed by examining the raw percentage scores on the Yes/Maybe/No and Circle tasks for all four students. The third graders and ninth graders were given exactly the same items on the tests of the animal defense topic, so we can compare performance across ages and abilities; no comparisons can be made on the Vocabulary measure because different words were used at each grade level.

[Insert Table 13 about here.]

The actual range of scores for these four students on the Yes/Maybe/No task is 73.33 to 86.67. The low-interview third-grade student received the highest score; she outperformed the two ninth-grade students as well as the high-interview third-grade student. Assessing these four students' topical knowledge with the Yes/Maybe/No task would lead us to conclude that the low-interview third-grade student knew more about animal defenses than the other three students. The range of scores on the Circle task is 69.05 to 78.57. On the Circle task, the high-interview ninth grader outperformed the others, but not by much (see Table 14).

[Insert Table 14 about here.]

The comparable performance of students across grade levels and across interview performance levels on identical Yes/Maybe/No and Circle tasks suggests that these measures are not discriminating

adequately for different levels of knowledge about the topic of animal defenses. Alternatively, it may be that they do discriminate, but not along the same lines as the interview. What is clear, however, is that different measures lead us to draw different conclusions about these students' topical knowledge.

Discussion

We began our work with the expectation that paper and pencil measures differed with respect to the degree that they served as good surrogate measures of topical knowledge. We expected to find evidence to support the construct and concurrent validity of one or more of our three paper and pencil measures; the best measures of topical knowledge, we reasoned, would be those that most closely mirrored the interview. The results were not straightforward. Instead of finding a *best* measure, we found that each measure contributed a different perspective on an individual's knowledge for a topic.

The trends point toward a complex model of topical knowledge--one that is comprised of two different spheres of knowledge--personal and communal. Often the overlap between the two is high; when it is, we tend to acknowledge this fact with expressions like, "We share a common ground." or "Everyone is starting from the same point." For a given topic, especially in school settings, there is likely to be some common knowledge that everyone possesses. For any given individual, there is almost always some personal knowledge--call it idiosyncratic--that is shared with few, if any, other members of the "community." There is also likely to be knowledge possessed by other members of the community but not that individual (Figure 4 provides a schematic representation of these relations).

[Insert Figure 4 about here.]

The situation is further complicated by the fact that not all common knowledge is equally salient and accessible, at least for any given individual. When people are asked to tell what they know about a topic (a very open-ended recall task), they omit many concepts that they will later recognize as highly related to the topic, noting afterward that those ideas just did not "occur to them" at the time. There are several possible explanations for this phenomenon.

One way to characterize these differences in the area of common knowledge is to invoke the distinction between *acquaintanceship* and *ownership* of concepts (Beck, Perfetti, & McKeown, 1982; Pearson, 1985). When we "own" an idea, we can recall it swiftly and easily, almost at will. But when we are acquainted with an idea, it comes to us only with prompting; we have to be reminded of it before we can share our knowledge about it. For example, one individual may speak at length about "camouflage" during an interview on the topic of animal defenses; in so doing she demonstrates *ownership* of that concept (this is represented by the striped area in Figure 4). A second individual may fail to mention the concept in an interview yet choose the correct meaning for camouflage on a Vocabulary test, thus demonstrating *acquaintanceship* with the concept (this is represented by the shaded area in Figure 4).

Our three pencil and paper measures of topical knowledge reflect aspects of knowledge more likely to be held by many (communal knowledge). Each measure gave us a somewhat different view of an individual's grasp of communal knowledge than was revealed by the interview. Some of the items included on the paper and pencil measures (at least those that were correctly answered) prodded the students' memories and allowed them to show knowledge they were acquainted with but did not own; hence they could not (or perhaps did not want to) share that knowledge voluntarily in an interview. Other items, those a given individual missed, came from that part of the communal sphere of knowledge that lay outside her personal knowledge.

It is possible that our ownership-acquaintanceship explanation oversimplifies the situation. Perhaps the differences we found between concepts recalled voluntarily in interviews versus those recognized on the paper and pencil tasks might stem from the way in which the information is organized and stored in memory. Voss (1984) found individuals, especially high-knowledge individuals, were able to identify concepts on a completion test that they did not give during free recall. He suggested an information

storage explanation: High-knowledge individuals may encode information in a hierarchical manner, making the relatively less important information harder to gain access to in a free recall situation. So what emerges in a recognition task is not so much knowledge that is "less well owned" as it is knowledge more deeply buried in semantic structures. Therefore, high-knowledge individuals may be more likely to access top-level knowledge, or more complex concepts, than lower level, more basic, common concepts. Their shared knowledge appears more personal and unique than communal.

In a related explanation, we might consider that measures of topical knowledge should vary across domains, reflecting the unique features of each domain's topical terrain. Spiro and his colleagues (Feltovich, Coulson, & Spiro, in press; Spiro, Vispoel, Schmitz, Samarapungavan, & Boerger, 1987) have argued that the structure of domain-specific knowledge falls along a continuum ranging from well-defined and well-structured to ill-defined and ill-structured. For example, human blood circulation is a topic that falls toward the well-defined and well-structured end of the continuum. It is fairly well circumscribed. The range of ideas and concepts that are considered relevant to the topic is fairly narrow. A discussion of the topic would include definitions of the applicable terminology as well as a description of the way the system functions. Creative interpretations of the information are less acceptable in a discussion of this topic. On the other hand, petsitting is a topic that falls near the other end of the continuum. This topic is not well circumscribed. The possible range of events and relevant concepts that could be included is very broad, which leads to the acceptance of differing notions of what constitutes relevance. This might suggest construction of topic familiarity measures that are responsive to the structures, patterns, and ways of thinking about particular knowledge domains. Thus, one possible explanation for the indeterminacy of our results in this study is that some formats we used were more sensitive in assessing the structures of one knowledge domain than were other formats.

Yet another explanation must be considered. It is possible that information from interviews was affected by the fact that they are socially mediated experiences that are affected by the perspective or focus the person chooses to take or perceives the interviewer to intend, as well as the motivation of the person being interviewed. In this kind of setting, individuals may consciously or unconsciously edit their public comments by deciding which aspects of their knowledge are relevant to the situation, what information the interviewer must already possess, what the interviewer really wants to hear and what information might be risky to share. In this explanation, the information we obtain from an interview may only tap the surface of an individual's knowledge, giving us a "tip of the iceberg" view.

Our design and data analysis procedures did not permit us to evaluate which of these explanations best fit our data. Regardless of how one accounts for the mismatches, what is important to note is that conclusions that might be drawn about an individual's knowledge are conditioned by the processes inherent in the measures we use and the contexts in which those measures are administered. Clearly, this is an area that needs to be explored by those who are interested in constructing valid measures of topical knowledge. Knowledge is a complex construct, and it may require complex assessment strategies; various measures may tap into very different aspects of this collage called knowledge. The assessment model we are now proposing is far more complex than our original model.

We began with the assumption that there exists a body of knowledge about a given topic that could be tapped fairly accurately and completely using an ideal measure, which we assumed to be an interview. We were interested in determining which paper and pencil measure most closely approximated the interview. Where we searched for simplicity we found complexity. Not only is there no one ideal method for ascertaining a person's knowledge about a topic, but different methods contribute unique information and portray different pictures of that individual's expertise. Paper and pencil measures serve the function of reminding students of information they know about the topic. Interviews give them the opportunity to share knowledge they have that is relevant to them personally, but not necessarily widely shared. And these two kinds of knowledge may or may not overlap.

What we have demonstrated here has far-reaching implications for the interpretation of past research and for future investigations into the connection between topical knowledge and comprehension. By

virtue of the design and content of a test of topical knowledge, the test constructor plays a critical role in predetermining and defining the structure and scope of topical knowledge deemed to be important. The test constructor defines the construct and ultimately defines its relation to other variables. It would appear that past efforts to measure topical familiarity have tried to simplify and "reduce" (Spiro, Vispoel, Schmitz, Samarapungavan, & Boerger, 1987) a construct that cannot be easily reduced; in brief, the integrity of the concept and the measures has been short-changed, and it is possible the field has been misled by the conclusions of the research.

Further research may help us gain an understanding of the multiple dimensions of topical knowledge and how these dimensions serve as predictors of comprehension. For example, it may be that we would have a different understanding of the topical knowledge/comprehension relationship if research addressed the differences between knowledge needed for comprehension of narrative texts versus expository texts, or if more research considered how knowledge in various domains is structured and stored. Because differences do exist, all types of knowledge should logically be considered as candidate predictors of comprehension. We need research to determine the contexts in which these types of knowledge are beneficial and whether various types of knowledge are generalizable across different reading tasks and contexts. The work here provides a beginning; now it must be applied to the topical knowledge/comprehension link. The next step is to look at which approaches to measuring topical knowledge best explain variation in the comprehension of different types of texts.

References

Alvermann, D. E., & Hynd, C. R. (1987). *Overcoming misconceptions in science: An on-line study of prior knowledge activation*. Paper presented at the annual meeting of the National Reading Conference, St. Petersburg, FL.

Alvermann, D. E., Smith, L. C., & Readence, J. E. (1985). Prior knowledge activation and the comprehension of compatible and incompatible text. *Reading Research Quarterly, 20*, 420-436.

Anderson, C. W., & Smith, E. L. (1984). Children's preconceptions and content-area textbooks. In G. G. Duffy, L. R. Roehler, & J. Mason (Eds.), *Comprehension instruction: Perspectives and suggestions* (pp.187-201). New York: Longman.

Anderson, R. C., & Freebody, P. (1983). Reading comprehension and the assessment and acquisition of word knowledge. In B. Hutson (Ed.), *Advances in reading/language research* (pp. 231-256). Greenwich, CT: JAI Press.

Anderson, R. C., Reynolds, D. D., Schallert, D. L., & Goetz, E. T. (1977). Frameworks for comprehending discourse. *American Educational Research Journal, 14*, 367-381.

Armbruster, B. B., & Ostertag, J. (1988). *Questions in elementary science and social studies textbooks* (Tech. Rep. No. 463). Urbana-Champaign: University of Illinois, Center for the Study of Reading.

Baldwin, R. S., Peleg-Bruckner, Z., & McClintock, A. H. (1985). Effects of topic interest and prior knowledge on reading comprehension. *Reading Research Quarterly, 20*, 497-505.

Beck, I. L., Perfetti, C. A., & McKeown, M. G. (1982). Effects of long-term vocabulary instruction on lexical access and reading comprehension. *Journal of Educational Psychology, 74*, 506-521.

Bransford, J. D., & Johnson, M. K. (1973). Consideration of some problems of comprehension. In W. G. Chase (Ed.), *Visual information processing*. New York: Academic Press.

Burgess, R. G. (1984). *In the field: An introduction to field research*. London: George Allen & UnWin.

Callahan, D., & Drum, P. A. (1984). Reading ability and prior knowledge as predictors of eleven and twelve year olds' text comprehension. *Reading Psychology, 5*, 145-154.

Chiesi, H., Spilich, G., & Voss, J. F. (1979). Acquisition of domain-related information in relation to high and low domain knowledge. *Journal of Verbal Learning and Verbal Behavior, 18*, 257-273.

Chou Hare, V. (1982). Preassessment of topical knowledge: A validation and an extension. *Journal of Reading Behavior, 14*, 77-85.

Chou Hare, V., & Devine, D. A. (1983). Topical knowledge and topical interest predictors of listening comprehension. *Journal of Educational Research, 76*, 157-160.

Comprehensive Test of Basic Skills (1983). Monterey, CA: CTB/McGraw-Hill.

Davey, B., & Kapinus, B. A. (1985). Prior knowledge and recall of unfamiliar information: Reader and text factors. *Journal of Educational Research, 78*, 147-151.

Domaracki, J. (1984). *The relationship of the amount and structure of children's prior knowledge to reading comprehension of expository prose*. Unpublished doctoral dissertation, State University of New York at Albany.

Dooling, D. J., & Lachman, R. (1971). Effects of comprehension on retention of words in episodic memory. *Journal of Experimental Psychology, 88*, 216-222.

Feltovich, P. G., Coulson, R. L., & Spiro, R. J. (in press). The nature of conceptual understanding in biomedicine: The deep structure of complex ideas and the development of misc[illegible]eptions. In D. Evans & V. Patel (Eds.), *The cognitive sciences in medicine*. Cambridge, MA: The MIT Press.

Finch, J. (1984). It's great to have someone to talk to: The ethics and politics of interviewing women. In C. Bell & H. Roberts (Eds.), *Social researching: Policies, problems, and practice* (pp. 70-87). London: Routledge & Kegan Paul.

Foertsch, M., & Pearson, P. D. (1987). *Reading assessment in basal reading series and standardized tests*. Paper presented at the annual meeting of the National Reading Conference, St. Petersburg, FL.

Gordon, R. L. (1980). *Interviewing: Strategies, techniques, and tactics*. Homewood, IL: The Dorsey Press.

Graves, M. F., Cooke, C. L., & LaBerge, M. J. (1983). Effects of previewing difficult short stories on low ability junior high school students' comprehension, recall, and attitudes. *Reading Research Quarterly, 18*, 262-276.

Holmes, B. C. (1983). The effects of prior knowledge on the question answering of good and poor readers. *Journal of Reading Behavior, 15*, 1-18.

Holmes, B. C. & Roser, N. L. (1987). Five ways to assess readers' prior knowledge. *The Reading Teacher, 40*, 646-649.

Johnson, D. D. & Pearson, P. D. (1978). *Teaching reading vocabulary*. New York: Holt, Rinehart & Winston.

Johnson, D. D. & Pearson, P. D. (1984). *Teaching reading vocabulary* (2nd Ed.). New York: Holt, Rinehart, & Winston.

Johnston, P. (1984a). Assessment in reading. In P. D. Pearson (Ed.), *Handbook of reading research* (pp. 147-182). New York: Longman.

Johnston, P. (1984b). Prior knowledge and reading comprehension test bias. *Reading Research Quarterly, 19*, 219-239.

Langer, J. A. (1980). Relation between levels of prior knowledge and the organization of recall. In M. L. Kamil & A. J. Moe (Eds.), Perspectives in reading research and instruction. *Twenty-ninth yearbook of the National Reading Conference* (pp. 28-33). Washington, DC: National Reading Conference.

Langer, J. A. (1984). Examining background knowledge and text comprehension. *Reading Research Quarterly, 29*, 468-481.

Langer, J. A., & Nicolich, M. (1981). Prior knowledge and its relationship to comprehension. *Journal of Reading Behavior, 13*, 373-379.

Lipson, M. Y. (1982). Learning new information from text: The role of prior knowledge and reading ability. *Journal of Reading Behavior, 14*, 243-261.

Mandler, J. M. (1984). *Stories, scripts, and scenes: Aspects of schema theory*. Hillsdale, NJ: Erlbaum.

Maria, K., & Blustein, E. (1986). *Refuting misconceptions: Its effect on middle grade children's comprehension*. Paper presented at the annual meeting of the National Reading Conference, Austin, TX.

Marr, M. B., & Gormley, K. (1982). Children's recall of familiar and unfamiliar text. *Reading Research Quarterly, 18*, 89-104.

Messick, S. (1981). Evidence and ethics in the evaluation of tests. *Educational Researcher, 10*, 9-20.

Oakley, A. (1981). Interviewing women: A contradiction in terms. In H. Roberts (Ed.), *Doing feminist research* (pp. 30-61). London: Routledge & Kegan Paul.

Olson, M. W. (1985). Text type and reader ability: The effects of paraphrase and text-based inference questions. *Journal of Reading Behavior, 17*, 199-214.

Pearson, P. D. (1985). Changing the face of reading comprehension instruction. *Ginn Occasional Papers*, (Number 21). Columbus, OH: Ginn & Company.

Pichert, J. W., & Anderson, R. C. (1977). Taking different perspectives on a story. *Educational Psychology, 69*, 309-315.

Spilich, G. J., Vesonder, G. T., Chiesi, H. L., & Voss, J. F. (1979). Text processing of domain-related information for individuals with high and low domain knowledge. *Journal of Verbal Learning and Verbal Behavior, 18*, 275-290.

Spiro, R. J., Feltovich, P. J., Coulson, R. L., & Anderson, D. (in press). Multiple analogies for complex concepts: Antidotes for analogy-induced misconception in advanced knowledge acquisition. In S. Vosniadou & A. Ortony (Eds.), *Similarity and analogical reasoning*. Cambridge: Cambridge University Press.

Spiro, R. J., Vispoel, W. L., Schmitz, J. L., Samarapungavan, A., & Boerger, A. E. (1987). Knowledge acquisition for application: Cognitive flexibility and transfer in complex knowledge domains. In B. C. Britton & S. Glynn (Eds.), *Executive control processes*. Hillsdale, NJ: Erlbaum

Stevens, K.C. (1980). The effect of background knowledge on the reading comprehension of ninth graders. *Journal of Reading Behavior, 12*, 151-154.

Valencia, S. V. & Pearson, P. D. (1987). *Reading assessment initiative in the state of Illinois--1985-1986*. Springfield: Illinois State Board of Education.

Voss, J. F. (1984). On learning and learning from text. In H. Mandl, N. L. Stein, & T. Trabasso (Eds.), *Learning and comprehension of text*. Hillsdale, NJ: Erlbaum

Webb, S., & Webb B. (1932). *Methods of social study*. London: Longman Green.

Table 1

Topical Knowledge Study Matrix

Study	Grade of Subjects	Type of Text	Type of PK	Type of Measure	When Measure Administered	Type of Scoring	Validation Attempted	Assess Interest
Alvermann, Smith, & Readence (1980)	6	E	Content	Multiple-Choice Recognition Questioning	Before Reading	Quantitative	No	No
Callahan & Drum (1984)	5 & 6	E	Content	Written Free Word Association	Before Reading	Quantitative	No	No
Chiesi, Spilich, & Voss (1979)	college	Semi-N	Content	Completion Questions	Before Reading	Quantitative	No	No
Chou Hare (1982)	6	E	Content and Structure	Written Free Word Association	Before Reading	Quantitative and Qualitative -organization	Indirectly	No
Chou Hare & Devine (1983)	1	N	Content	Multiple-Choice Recognition Questioning	Before & After Reading	Quantitative	Indirectly	Yes
Davey & Kapinus (1985)	8	E	Content	Structured and Direct Questioning	Before Reading	Quantitative	No	No
Domaracki (1984)	6 & 7	E	Content and Structure	Vocabulary Tests & Structured Word Asan.	Before Reading	Quantitative and Qualitative -organization	Indirectly	No
Holmes (1983)	5	E	Content	Structured and Direct Questioning	Before Reading	Quantitative and Qualitative -ideas	No	No

Table 1 (Continued)

Study	Grade of Subjects	Type of Text	Type of PK	Type of Measure	When Measure Administered	Type of Scoring	Validation Attempted	Assess Interest
Holmes & Roser (1987)	--	--	Content	First Types 1) 3-Oral Free Recall (Content & Personal) 2) Oral Structured & Direct Questioning 3) Oral Word Association 4) Multiple-Choice Recognition	Before Reading	Quantitative	Indirectly	
Johnston (1984)	8	II	Content	Vocabulary Test	Before Reading	Quantitative	Indirectly	No
Joseph & Dwyer (1984)	10	II	Content	Multiple-Choice Recognition Questioning	Before Reading	Quantitative	No	No
Langer (1980)	12	II	Content and Structure	Written Free Word Association	Before Reading	Qualitative -organization	No	No
Langer (1984)	6	II	Content and Structure	Written Free Word Association	Before Reading	Qualitative -organization	No	No
Langer & Nicolich (1981)	12	II	Content and Structure	Written Free Word Association	Before Reading	Qualitative -organization	No	No

Table 1 (Continued)

Study	Grade of Subjects	Type of Text	Type of PK	Type of Measure	When Measure Administered	Type of Scoring	Validation Attempted	Assess Interest
Lipson (1982)	3	E	Content	Multiple-Choice Recognition Questioning	Before Reading	Quantitative	No	No
Marr & Gormley (1982)	4	E	Content	Structured and Direct Questioning	Before Reading	Quantitative	No	No
Pearson, Hansen & Hansen & Gordon (1979)	2	E	Content	Structured and Direct Questioning	Before Reading	Quantitative	No	No
Spilich, Vesonder, Chiesi, & Voss (1979)	college	E	Content	Completion Questions	Before Reading	Quantitative	No	No
Stevens (1980)	9	E	Content	Multiple-Choice Recognition Questioning	Before Reading	Quantitative	No	No

Table 2

Means and Standard Deviations of all Formats for Grade 3

	Passage 3			Passage 4			Passage 5			Passage 6		
	Mean	Std	Range	Mean	Std	Range	Mean	Std	Range	Mean	Std	Range
Vocabulary	87.56	11.79	50-100	72.93	16.01	20-100	81.95	14.70	20-100	76.10	16.56	40-100
Yes/Maybe/No	66.89	11.69	40-87	74.40	13.33	43-97	58.04	13.09	33-87	72.32	13.05	33-100
Circle	67.42	13.61	33-100	----------			56.85	6.64	42-74	----------		
Interview	30.34	13.37	8-55	26.29	12.10	15-46	28.12	11.27	7-68	26.54	7.69	10-67

Passages

1 = Human Circulation (Expository)
2 = Spy (Narrative)
3 = Animal Defenses (Expository)
4 = Petsitter (Narrative)
5 = Plants and People (Expository)
6 = New Boy (Narrative)

Table 3

Means and Standard Deviations of all Formats for Grade 9

	Passage 3			Passage 4			Passage 1			Passage 2		
	Mean	Std	Range	Mean	Std	Range	Mean	Std	Range	Mean	Std	Range
Vocabulary	73.20	16.20	40-100	56.13	18.38	30-90	46.45	24.70	10-100	75.48	13.87	40-100
Yes/Maybe/No	67.20	12.50	40-87	78.60	13.10	47-100	66.14	13.38	33-87	77.85	10.17	50-93
Circle	70.60	12.31	38-88	----------			68.66	13.41	38-93	----------		
Interview	34.94	13.22	16-69	29.81	10.39	12-62	18.48	10.09	5-50	20.58	7.06	9-38

Passages

1 = Human Circulation (Expository)
2 = Spy (Narrative)
3 = Animal Defenses (Expository)
4 = Petsitter (Narrative)
5 = Plants and People (Expository)
6 = New Boy (Narrative)

Table 4

Correlations Between Interview Scores and Scores on Paper and Pencil Tests by Passage

Grade Three

	Passage			
	3	4	5	6
Vocabulary	.42**	.32*	.52**	.29*
Yes/Maybe/No	.28	.13	.42**	-.14
Circle	.46**		***	

Grade Nine

	Passage			
	3	4	1	2
Vocabulary	.28	.22	.71**	-.02
Yes/Maybe/No	.25	.14	.33	-.19
Circle	-.01		.11	

*$p < .05$

**$p < .01$

***Unable to calculate

Passages
1 = Human Circulation (Expository)
2 = Spy (Narrative)
3 = Animal Defenses (Expository)
4 = Petsitter (Narrative)
5 = Plants and People (Expository)
6 = New Boy (Narrative)

Table 5

Regression of Paper and Pencil Tests on the Interview Scores by Passage and Grade Level

(Amount of variance in interview scores accounted for by all the paper and pencil measures.)

Grade Three

	Passage			
	3	4	5	6
R^2	.31*	.10	.35*	.14

Grade Nine

	Passage			
	3	4	1	2
R^2	.11	.06	.58*	.09

*$p < .05$

Passages
1 = Human Circulation (Expository)
2 = Spy (Narrative)
3 = Animal Defenses (Expository)
4 = Petsitter (Narrative)
5 = Plants and People (Expository)
6 = New Boy (Narrative)

Table 6

Reliability Coefficients for Paper and Pencil Tests (Cronbach's Alpha)

Grade Three

	Passage			
	3	4	5	6
Vocabulary	.51	.58	.79	.52
Yes/Maybe/No	.36	.47	.51	.65
Circle	.56		***	

Grade Nine

	Passage			
	3	4	1	2
Vocabulary	.58	.50	.68	.49
Yes/Maybe/No	.48	.63	.59	.45
Circle	.73		.57	

***Unable to calculate

Passages
1 = Human Circulation (Expository)
2 = Spy (Narrative)
3 = Animal Defenses (Expository)
4 = Petsitter (Narrative)
5 = Plants and People (Expository)
6 = New Boy (Narrative)

Table 7

Correlations Between Interview Scores and Scores on Paper and Pencil Tests by Passage, after Correcting for Attenuation

Grade Three

	Passage			
	3	4	5	6
Vocabulary	.58*	.42*	.58*	.40*
Yes/Maybe/No	.47*	.19	.59*	-.17
Circle	.61*		***	

Grade Nine

	Passage			
	3	4	1	2
Vocabulary	.37*	.31*	.86*	-.03
Yes/Maybe/No	.36*	.18	.43*	.21
Circle	-.01		.14	

*$p < .05$

***Unable to calculate

Passages
1 = Human Circulation (Expository)
2 = Spy (Narrative)
3 = Animal Defenses (Expository)
4 = Petsitter (Narrative)
5 = Plants and People (Expository)
6 = New Boy (Narrative)

Table 8

Mean (Standard Deviations) Consistency of Performance Between the Interview and Various Paper and Pencil Measures for Grade 3 Narrative

	Measures		
	Interview	Vocabulary	Yes/Maybe/No
	Passage 4 (Petsitter)		
Interview		92.77 (19.27)[1]	75.16 (35.63)[2]
Vocabulary	42.04 (20.19)[3]		
Yes/Maybe/No	19.87 (19.99)[4]		
	Passage 6 (New Boy)		
Interview		75.27 (32.29)[1]	74.10 (26.80)[2]
Vocabulary	40.62 (23.51)[3]		
Yes/Maybe/No	47.48 (26.84)[4]		

1 = Percentage of ideas given in the Interview that corresponded to items that were correctly answered on the Vocabulary test.

2 = Percentage of ideas given in the Interview that corresponded to items that were correctly answered on the Yes/Maybe/No test.

3 = Of all the items answered correctly on the Vocabulary test, the percentage that corresponded to ideas mentioned during the Interview.

4 = Of all the items answered correctly on the Yes/Maybe/No test, the percentage that corresponded to ideas mentioned during the Interview.

Table 9

Mean (Standard Deviations) Consistency of Performance Between the Interview and Various Paper and Pencil Measures for Grade 3 Expository Topics

	Measures			
	Interview	Vocabulary	Yes/Maybe/No	Circle
		Passage 3 (Animal Defenses)		
Interview		97.62 (10.02)[1]	90.56 (22.50)[2]	74.08 (26.99)[3]
Vocabulary	24.00 (19.04)[4]			
Yes/Maybe/No	26.17 (22.17)[5]			
Circle	28.06 (17.31)[6]			
		Passage 5 (Plants and People)		
Interview		94.00 (21.98)[1]	71.62 (29.18)[2]	56.17 (29.03)[3]
Vocabulary	18.34 (17.74)[4]			
Yes/Maybe/No	45.44 (26.27)[5]			
Circle	38.30 (27.27)[6]			

1 = Percentage of ideas given in the Interview that corresponded to items that were correctly answered on the Vocabulary test.

2 = Percentage of ideas given in the Interview that corresponded to items that were correctly answered on the Yes/Maybe/No test.

3 = Percentage of ideas given in the Interview that corresponded to items that were correctly answered on the Circle test.

4 = Of all the items answered correctly on the Vocabulary test, the percentage that corresponded to ideas mentioned during the Interview.

5 = Of all the items answered correctly on the Yes/Maybe/No test, the percentage that corresponded to ideas mentioned during the Interview.

6 = Of all the items answered correctly on the Circle test, the percentage that corresponded to ideas mentioned during the Interview.

Table 10

Mean (Standard Deviations) Consistency of Performance Between the Interview and Various Paper and Pencil Measures for Grade 9 Narrative Topics

	Measures		
	Interview	Vocabulary	Yes/Maybe/No
	Passage 4 (Petsitter)		
Interview		50.63 (28.54)[1]	79.81 (34.25)[2]
Vocabulary	51.98 (24.25)[3]		
Yes/Maybe/No	24.97 (18.76)[4]		
	Passage 2 (Spy)		
Interview		80.86 (32.94)[1]	87.10 (18.43)[2]
Vocabulary	23.10 (14.25)[3]		
Yes/Maybe/No	42.53 (17.36)[4]		

1 = Percentage of ideas given in the Interview that corresponded to items that were correctly answered on the Vocabulary test.

2 = Percentage of ideas given in the Interview that corresponded to items that were correctly answered on the Yes/Maybe/No test.

3 = Of all the items answered correctly on the Vocabulary test, the percentage that corresponded to ideas mentioned during the Interview.

4 = Of all the items answered correctly on the Yes/Maybe/No test, the percentage that corresponded to ideas mentioned during the Interview.

Table 11

Mean (Standard Deviations) Consistency of Performance Between the Interview and Various Paper and Pencil Measures for Grade 9 Expository Topics

	Measures			
	Interview	Vocabulary	Yes/Maybe/No	Circle
		Passage 3 (Animal Defenses)		
Interview		91.05 (25.50)[1]	75.11 (37.61)[2]	76.96 (23.01)[3]
Vocabulary	32.17 (15.31)[4]			
Yes/Maybe/No	31.09 (26.25)[5]			
Circle	29.07 (11.63)[6]			
		Passage 1 (Human Circulation)		
Interview		65.28 (47.85)[1]	50.00 (50.00)[2]	66.67 (44.41)[3]
Vocabulary	28.01 (36.54)[4]			
Yes/Maybe/No	14.29 (29.99)[5]			
Circle	16.06 (23.17)[6]			

1 = Percentage of ideas given in the Interview that corresponded to items that were correctly answered on the Vocabulary test.

2 = Percentage of ideas given in the Interview that corresponded to items that were correctly answered on the Yes/Maybe/No test.

3 = Percentage of ideas given in the Interview that corresponded to items that were correctly answered on the Circle test.

4 = Of all the items answered correctly on the Vocabulary test, the percentage that corresponded to ideas mentioned during the Interview.

5 = Of all the items answered correctly on the Yes/Maybe/No test, the percentage that corresponded to ideas mentioned during the Interview.

6 = Of all the items answered correctly on the Circle test, the percentage that corresponded to ideas mentioned during the Interview.

Table 12

Mean Percent of Unique Information for Each Topical Knowledge Measure

	Grade 3			
Passages	3	4	5	6
Interview	74.18 (11.69)	50.74 (16.11)	71.79 (12.30)	62.26 (17.87)
Vocabulary	77.37 (17.57)	63.42 (20.42)	84.65 (15.13)	60.87 (23.37)
Circle	78.40 (12.47)	NA	73.28 (20.20)	NA
Yes/Maybe/No	80.34 (20.40)	82.53 (18.76)	69.18 (17.52)	61.47 (23.54)
	Grade 9			
Passages	1	2	3	4
Interview	87.23 (9.46)	58.11 (24.16)	69.50 (10.35)	53.91 (12.66)
Vocabulary	88.78 (14.10)	81.34 (12.93)	74.61 (12.48)	71.91 (13.09)
Circle	90.45 (15.28)	NA	76.38 (10.93)	NA
Yes/Maybe/No	96.51 (7.38)	68.18 (13.99)	79.76 (17.59)	77.73 (16.19)

Passages

1 = Human Circulation (Expository)
2 = Spy (Narrative)
3 = Animal Defenses (Expository)
4 = Petsitter (Narrative)
5 = Plants and People (Expository)
6 = New Boy (Narrative)

Table 13

Four Views of Four Students' Topical Knowledge on Animal Defenses[a]

Grade	Interview	Vocabulary	Yes/Maybe/No	Circle
3	38.52	52.07	63.89	51.42
3	75.18	43.58	53.53	56.79
9	37.94	48.01	52.24	48.74
9	70.47	48.01	49.57	52.61

[a]All scores are standardized within grade level with a mean of 50 and a standard deviation of 10.

Table 14

Four Students' Percentage Scores on the Yes/Maybe/No and Circle Measures[a]

	Yes/Maybe/No	Circle
Grade 3		
Low Interview	86.67	69.05
High Interview	73.33	76.20
Grade 9		
Low Interview	76.67	71.43
High Interview	73.33	78.57

[a]Scores range from 0 to 100 on both measures.

Figure 1. Continuum of Topical Knowledge Formats Ranging From Recognition to Recall.

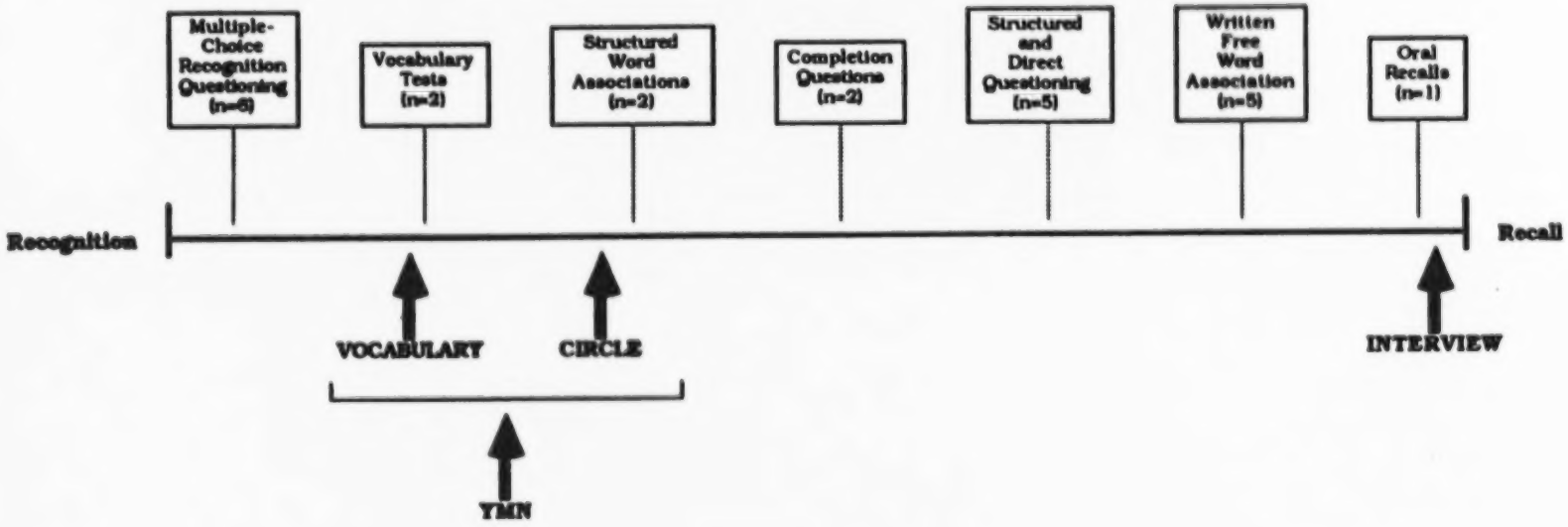
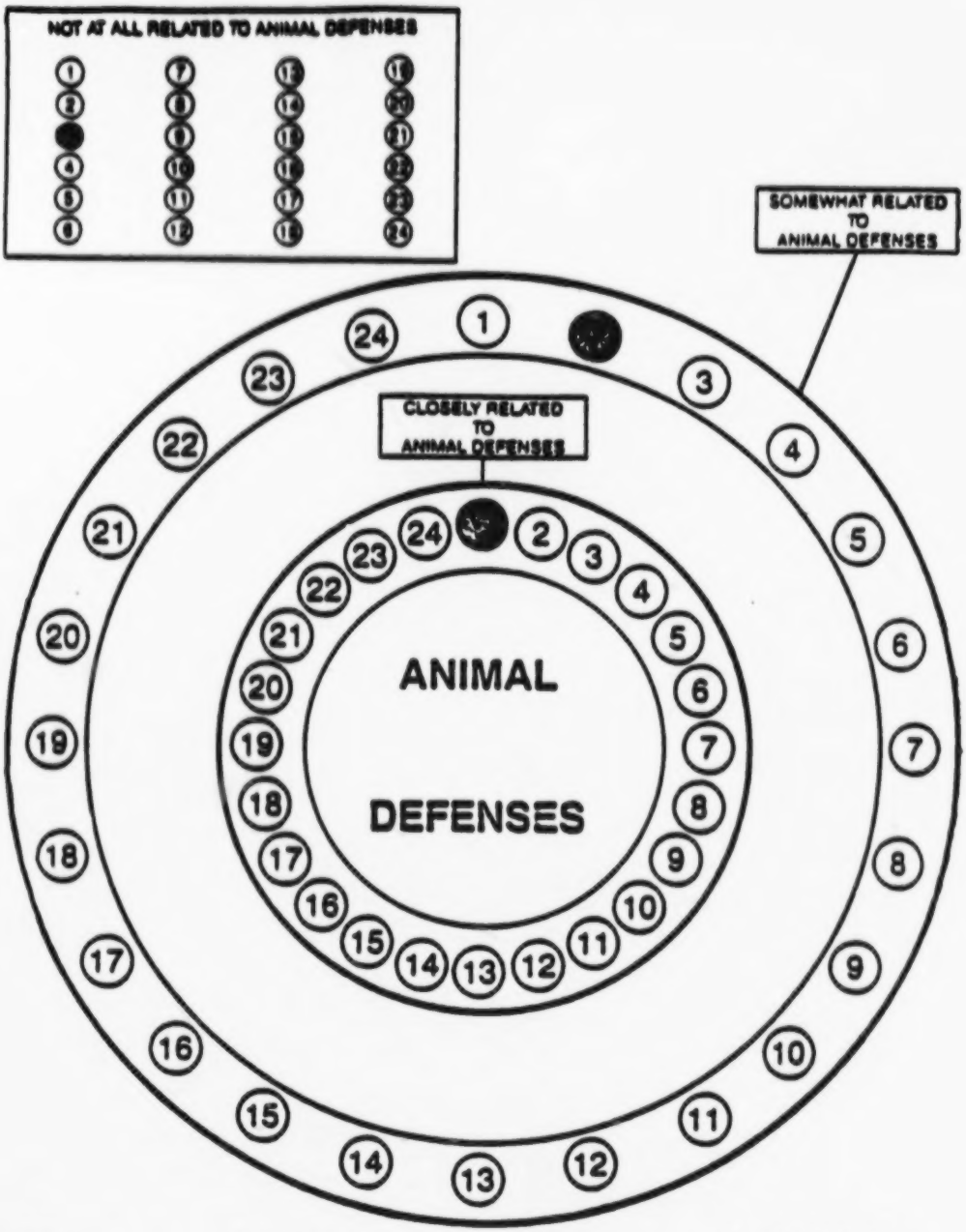


Figure 2. Circle Measure for Animal Defenses



1. color	7. venom	13. prey	19. graphology
2. nature	8. trapping	14. reptiles	20. territorial
3. cardboard	9. biology	15. attack	21. habitat
4. crater	10. erosion	16. omnivorous	22. camouflage
5. predator	11. influenza	17. scavenger	23. hydrogen
6. carnivorous	12. storm	18. speed	24. escape

Figure 3

Yes/Maybe/No Measure for Animal Defenses

Pretend that you are going to read a library or science book about how animals hunt other animals to get food and how the hunted animals try to protect themselves. Think about what you know about how animals hunt one another for food and about how the hunted animals defend themselves. Below you will find several ideas that you might or might not find in a book like this one. For each idea, decide whether or not you might find it in a book like this one. Then fill in the bubble that tells what you think. The first 3 have been done for you.

(YES) = Yes, I think it is very likely that the idea would be in an article about the topic.
(MAYBE) = Maybe the idea could be in the article.
(NO) = No, I don't think the idea would be in an article about the topic.

	YES	MAYBE	NO
a. Animals that hide in the forest	(Y)	(M)	(N)
b. How animals find water	(Y)	(M)	(N)
c. Learning how to ride a bicycle	(Y)	(M)	(N)
1. How animals feed their young	(Y)	(M)	(N)
2. Predators	(Y)	(M)	(N)
3. Clothing from animal skins	(Y)	(M)	(N)
4. Poisonous animals	(Y)	(M)	(N)
5. Building nests	(Y)	(M)	(N)
6. Camouflage	(Y)	(M)	(N)
7. Animal armour	(Y)	(M)	(N)
8. Carnivorous animals	(Y)	(M)	(N)
9. Scavengers	(Y)	(M)	(N)
10. Herbivorous animals	(Y)	(M)	(N)
11. Feeding zoo animals	(Y)	(M)	(N)
12. Chlorophyll in plants	(Y)	(M)	(N)
13. Animals that swim	(Y)	(M)	(N)
14. Caring for horses	(Y)	(M)	(N)
15. Bright colors of animals	(Y)	(M)	(N)

Figure 4. Schematic Representation of Topical Knowledge Relationships